

**Physical similarity or numerical representation counts in same-different, numerical comparison, physical comparison, and priming tasks?**

Li Zhang<sup>1</sup>, Ziqiang Xin<sup>2</sup>, Tingyong Feng<sup>1</sup>, Yinghe Chen<sup>3</sup>, Denes Szűcs<sup>4</sup>

<sup>1</sup> Faculty of Psychology, Southwest University, 400715, China

<sup>2</sup> Department of Psychology, School of Sociology and Psychology, Central University of Finance and Economics, Beijing 100081, China

<sup>3</sup> Institute of Developmental Psychology, School of Psychology, Beijing Normal University, Beijing 100875, China

<sup>4</sup> Department of Psychology, University of Cambridge, Cambridge, CB2 3EB, UK

**Acknowledgments**

This study was funded by the National Natural Science Foundation of China (Grant No. 31470996) and Major Project of the National Social Science Foundation of China (Grant No. 14ZDB160). Correspondence concerning this article should be addressed to Ziqiang Xin, [xinziqiang@sohu.com](mailto:xinziqiang@sohu.com) or Denes Szűcs, [ds377@cam.ac.uk](mailto:ds377@cam.ac.uk)

## **Abstract**

Recent studies have highlighted the fact that some tasks used to study symbolic number representations are confounded by judgments about physical similarity. Here, we investigated whether the contribution of physical similarity and numerical representation differed in the often used symbolic same-different, numerical comparison, physical comparison, and priming tasks. Experiment 1 showed that subjective physical similarity was the best predictor of participants' performance in the same-different task, regardless of simultaneous or sequential presentation. Furthermore, the contribution of subjective physical similarity was larger in a simultaneous presentation than in a sequential presentation. Experiment 2 showed that only numerical representation was involved in numerical comparison. Experiment 3 showed that both subjective physical similarity and numerical representation contributed to participants' physical comparison performance. Finally, only numerical representation contributed to participants' performance in a priming task as revealed by Experiment 4. Taken together, the contribution of physical similarity and numerical representation depends on task demands. Performance primarily seems to rely on numerical properties in tasks which require explicit quantitative comparison judgments (physical or numerical), while physical stimulus properties exert an effect in the same-different task.

**Key words:** physical similarity, numerical representation, same-different task, numerical comparison task, physical comparison task, priming task

## ***1 Introduction***

Humans primarily employ symbolic numbers, such as Arabic digits to convey numerical magnitudes. The processing of symbolic numbers has been studied by analyzing their magnitude relations. Important evidence comes from the finding that the numerical distance contributes to the reaction time of numerical processing. This finding is consistently observed in the same-different task (e.g., Dehaene & Akhavein, 1995; Duncan & McFarland, 1980; Van Opstal & Verguts, 2011), in the numerical comparison task (e.g., Moyer & Landauer, 1967), in the Stroop-like physical comparison task (e.g., Ganor-Stern & Tzelgov, 2008), and in the priming task (e.g., Reynvoet, Brysbaert, & Fias, 2002). However, several recent studies have argued that the physical similarity between Arabic numbers, other than the numerical distance, contributes to the numerical processing in the same-different task (e.g., Cohen, 2009; Defever, Sasanguie, Vandewaetere, & Reynvoet, 2012; Garcia-Orza, Perea, Abu Mallouh, & Carreiras, 2012; Wong & Szűcs, 2013). In this study, we further investigated the contribution of numerical distance and physical similarity by extending the original experimental tasks to novel presentation conditions and to major tasks used in the field of numerical cognition.

### ***1.1 Background***

It is widely thought that humans possess an internal semantic representation of magnitude that corresponds with symbolic numbers in the external environment (Domahs et al., 2012; Lyons & Ansari, 2008; Verguts & Fias, 2004; Young & Opfer, 2011; Zhang & Wang, 2005). Therefore, symbolic numbers naturally activate such

magnitude representations. In numerical cognition literature, a particularly well researched potential marker of such magnitude representations is the so-called numerical distance effect. This effect was originally observed in the comparison task (e.g., Moyer & Landauer, 1967). When subjects have to judge which of two magnitudes is numerically larger, close magnitudes (or numbers) are more difficult (slower and more error prone) to compare than distant magnitudes (e.g. comparing 1 vs. 5 is faster and more accurate than comparing 1 vs. 2). Furthermore, Moyer and Landauer showed that the numerical distance effect was described by the Welford (1960) function (henceforth abbreviated as  $N_W$ ):  $RT = a + k * \log [L / (L-S)]$ , where  $a$  and  $k$  are constants,  $L$  is the larger quantity and  $S$  is the smaller quantity. The distance effect in the comparison task has been repeated in several studies (e.g., Buckley & Gillman, 1974; Dehaene, 1989; Duncan & McFarland, 1980; Hinrichs, Yurko, & Hu, 1981; Moyer & Landauer, 1973; Shepard et al., 1975).

The distance effect which is often described as the absolute distance between the larger number and the smaller number (henceforth abbreviated as  $N_D$ ) also appears in the same-different task (e.g., Dehaene & Akhavein, 1995; Duncan & McFarland, 1980; Ganor-Stern & Tzelgov, 2008; Van Opstal & Verguts, 2011; Verguts & Van Opstal, 2005). In the task, subjects are required to judge whether two magnitudes are the “same” or “different”. Different responses are faster when the two numbers are numerically far apart rather than when they are close.

Even in the physical comparison task in which the numerical magnitude of numbers is irrelevant to the task requirement,  $N_D$  can be found (e.g., Cohen Kadosh, Henik, &

Rubinsten, 2008; Kaufmann, et al., 2005; Szűcs & Soltesz, 2007, 2008; Tzelgov et al., 1992; Zhou et al., 2008). In the physical comparison task, which is so-called numerical stroop task, participants have to decide which one of two simultaneously presented Arabic digits is larger than the other in physical size (Henik & Tzelgov, 1982). The relationship of numerical and physical size features of the digits may be neutral, congruent, or incongruent. In the congruent condition the numerically larger digit is physically larger than the other digit (e.g., 2 and 7). In the incongruent condition the numerically larger digit is physically smaller than the other (e.g., 2 and 7). In the neutral condition the digits do not differ on the task-irrelevant stimulus dimension (e.g., 2 and 7). Usually, performance is better in the congruent condition than in neutral and incongruent conditions, which suggests that the numerical magnitude of numbers is automatically represented. Although the numerical magnitudes of numbers are not intentionally processed, the distance effect can still be observed in this task (e.g., Cohen Kadosh, Henik, & Rubinsten, 2008; Kaufmann, et al., 2005; Szűcs & Soltesz, 2007, 2008; Tzelgov et al., 1992; Zhou et al., 2008), which confirms the activation of the numerical magnitude of numbers.

In addition,  $N_D$  is also observed in priming studies (Notebaert, Pesenti, Reynvoet, 2010; Perry & Lupker, 2012; Reynvoet, Brysbaert, & Fias, 2002; Reynvoet, De Smedt, Van den Bussche, 2009; Van Opstal, Gevers, De Moor, & Verguts, 2008). This effect is different from that shown in the above three tasks in that performances to target numbers are better when they are preceded by a numerically close prime than by a numerically far prime. For example, the digit “2” is named faster when it is

preceded by “3” than by “4” (Reynvoet, Brysbaert, & Fias, 2002).

To sum up, the distance effect consistently appears in all the above number tasks. This effect seems to be the primary factor controlling participants’ performance. However, in recent years several studies have identified an alternative perceptual factor to explain participants’ performance (e.g., Cohen, 2009; Cohen, Warren, & Goldhammer, 2013; Defever et al., 2012; Garcia-Orza et al., 2012; Wong & Szűcs, 2013). In the seminal study by Cohen (2009), participants were presented with single Arabic number between 1 and 9 and asked to decide whether the number presented was 5 or not 5. Participants’ average RTs were regressed on the Welford function ( $N_W$ ) and on the physical similarity function (henceforth abbreviated as  $P_C$ ). The formula for  $P_C$  was  $P_C = O/D$ , where  $O$  is the number of lines that the two digits share (i.e. overlapping lines), and  $D$  is the number of unshared lines between the two digits (i.e. difference). The results showed that both predictors contributed to the regression, but, more importantly, when both functions were included as predictors, only  $P_C$  was significant.

Garcia-Orza et al. (2012) also tested the contribution of physical similarity and numerical distance to participants’ performance in the 5-or-not-5 task. In their study, three groups of participants were chosen: Pakistani students, who were fluent only in the Persian version of Indian numbers; Jordanian students, who were fluent only in the Arabic version of Indian numbers; and Spanish students, who had no knowledge of either. In Experiment 1, Spanish and Pakistani participants completed a 5-or-not-5 task with Persian–Indian numbers. The results showed that when  $N_W$  and  $P_C$  were

entered simultaneously in a multiple regression,  $P_C$  was the only significant predictor. In Experiment 2, Spanish and Jordanian participants completed the 5-or-not-5 task with Arabic–Indian numbers. Similarly, researchers found that  $P_C$  was the only significant predictor of Jordanian students’ reaction time.

Defever et al. (2012) investigated the development of magnitude representations in children aged 5 to 11 using a symbolic same–different task. The physical similarity was defined as subjective visual similarity ( $P_S$ ) which was a subjective rating for every possible combination of Arabic digit pairs using a seven point scale (Campbell & Clark, 1988). The numerical distance was described as the absolute distance between the larger number and the smaller number ( $N_D$ ). The researchers found a main effect of  $P_S$  but no effect of  $N_D$  in children. Similarly, Wong and Szűcs (2013) explored whether reaction time in the symbolic same-different task is primarily driven by objective or subjective perceptual similarity, or by the numerical difference between the two digits. The regression analyses showed that subjective  $P_S$  was the best and only significant predictor of reaction time in children and adults.

Recently, Cohen, Warren, Blanc-Goldhammer (2013) have explored the role of physical similarity in the same-different and numerical comparison tasks with Arabic digits, written number words and auditory numbers words as stimuli. In Experiment 1, they assessed the relative influence of  $P_C$  and  $N_W$  in the same-different task. Two Arabic digits (digit–digit), two written number words (word–word), or an Arabic digit and a written number word (word–digit or digit–word) were presented one above the other. Results showed that in all conditions  $P_C$  was the primary controlling variable.

Experiment 2 was similar to Experiment 1 except that the numerical comparison task was used. Results showed that  $N_W$  was the primary controlling variable while  $P_C$  was also a significant predictor in three of four conditions. In Experiments 3 and 4, two Arabic digits (digit–digit) or an auditory number word and an Arabic digit (auditory–digit) were presented to subjects. Similarly, the  $P_C$  was the primary controlling variable in the same-different task while  $N_W$  was the primary controlling variable in the numerical comparison task. The authors claimed that they observed the cross-format physical similarity effect.

### ***1.2 The present study***

In this present study, we had several objectives. First, we sought to clarify whether objective or subjective physical similarity is the primary predictor of participants' performance in the same-different task as up to this point there have been mixed results regarding this question. Three studies mentioned above showed that objective physical similarity was the primary predictor of participants' performance (e.g., Cohen, 2009; Cohen, Warren, & Goldhammer, 2013; Garcia-Orza et al, 2012) while two studies (Defever et al, 2012; Wong & Szűcs, 2013) found that it is subjective physical similarity that best predicts participants' performance in the same-different task.

Second, we sought to investigate the effect of task demands on the role of physical similarity in the processing of Arabic numbers. Typically, the same-different task has been used by previous studies which have demonstrated the important role of physical similarity in the processing of Arabic numbers (e.g., Cohen, 2009; Defever et al.,



2012; Garcia-Orza et al., 2012; Wong & Szűcs, 2013). The 5-or-not-5 task is actually a special version of the same-different task where participants are required to judge whether the target number is the same with 5 or different from 5. The main problem with the same-different task is that this task can be completed simply by analyzing the surface perceptual structure of the presented numerical symbols when the two numbers were presented simultaneously. The important role of physical similarity may only be valid under specific tasks. It is unclear whether it can be generalized to other tasks, such as the numerical comparison task, the physical comparison task, and the priming task where the numerical distance effect can be usually observed. For the first time, we would test the contribution of physical similarity and numerical distance in physical comparison and priming tasks. In the study by Cohen et al. (2013), the role of physical similarity in the numerical comparison task has been explored; however, this study differed from it in two ways (Cohen et al, 2013). First, we would distinguish subjective from objective physical similarity. Second, we would examine the possible mechanism among the relationship of physical similarity, numerical distance and participants' performance. In the study by Cohen et al. (2013), the physical similarity was found to be another significant predictor of participants' performance in addition to the numerical distance in the numerical comparison task; however, it is not clear whether the effect of physical similarity on participants' performance is mediated by the numerical distance. This question is examined in our study.

The third aim of our study was to investigate the effect of presentation mode,

whether simultaneous or sequential, on the role of physical similarity in the processing of Arabic numbers. Some previous studies have revealed the important influence of presentation mode on two-digit number processing (Ganor-Stern, Pinhas, & Tzegov, 2009; Zhang & Wang, 2005; Zhou, Chen, Chen, & Dong, 2008). In the simultaneous presentation, there is evidence for parallel processing with the digits in both the units and decades decimal places. In contrast, in the sequential presentation, there is evidence for processing of mainly the decades place while no specific influence of the units place on task performance is observed. In this study, similarly to the study by Wong and Szűcs (2013), we would use the symbolic same-different task where participants were visually presented pairs of single-digit Arabic numbers and asked to judge whether they were the same or different. This task has an advantage over the frequently used 5-or-not-5 task in that it can include all possible combinations of Arabic numbers (Wong & Szűcs, 2013). However, for the first time we would test if the presentation mode in the same-different task influences the contribution of physical similarity in the processing of single-Arabic numbers. By doing so, we can test the general validity of the important role of physical similarity in the processing of Arabic numbers under different presentation conditions.

Finally, this study also aimed to explore whether symbolic numbers were represented on a linear or a log scale under different task demands. When symbolic numbers are represented on a linear scale, the absolute distance between two numbers ( $N_D$ ) will predict participants' performance; when symbolic numbers are represented on a log scale, the Welford function ( $N_W$ ) will predict participants' performance. Most

of previous studies on symbolic number processing used absolute distance as a metric of the distance effect, even in the numerical comparison task where participants have been shown to rely more on logarithmic representations (Hinrichs, Yurko, & Hu, 1981; Moyer & Landauer, 1967). Using a neural simulation method, Verguts and Fias (2004) indicated that both linear and log scaling produced a good fit for symbolic stimuli. However, it is unclear whether the linear or log representation of symbolic number is dependent on task demands. In this study, therefore, we addressed whether linear or log scaling produced a better fit for symbolic stimuli under different task demands.

We have conducted four experiments in order to address the above objectives. In Experiment 1, we examined whether objective or subjective physical similarity could best predict participants' performance in the same-different task and whether the prediction was affected by the mode of presentation. In Experiments 2, 3, and 4 we examined the contribution of physical similarity (objective and subjective) and numerical distance in numerical comparison, physical comparison and priming tasks respectively.

There were four hypotheses. First, we hypothesized that the subjective physical similarity would be the best predictor of participants' performance in the same-different task. This is based on the finding that the subjective physical similarity was the best and only significant predictor of reaction time in adults and children when the effect of both subjective and objective physical similarity on reaction time was assessed (Wong & Szűcs, 2013). Second, we hypothesized that the contribution of subjective physical similarity would be larger in a simultaneous presentation than in a

sequential presentation. In a sequential presentation where one number has to be kept in memory, the perceptual influence may be reduced. Zhang and Wang (2005) have argued that in a sequential presentation the comparison is between an external representation of the number currently presented and an internal representation in memory of the previously presented number. Such a situation requires a translation of the representations into a common form before the numbers can be compared. In line with the rationale proposed by Zhang and Wang (2005), in a sequential presentation, participants in our study would be expected to use a common internal magnitude representation to complete the same-different task. Third, based on the study by Cohen et al. (2013), we hypothesized that the important role of subjective physical similarity should not be valid in numerical comparison task. As for the contribution of physical similarity in physical comparison and priming tasks, we did not have clear hypotheses since no previous study has addressed this question. Finally, the log scaling of symbolic number should produce a better fit in the numerical comparison task as revealed by some studies (e.g., Hinrichs, Yurko, & Hu, 1981; Moyer & Landauer, 1967). Given no relevant studies, we did not have any clear hypothesis regarding the question that a linear or a log scale had a better fit in the other three tasks.

## ***2 Experiment 1: same-different task***

In Experiment 1, we assessed the relative influence of physical similarity and numerical distance in simultaneous and sequential same-different tasks.

### ***2.1 Methods***

### **2.1.1 Participants**

Twenty undergraduates were recruited from Southwest University, China. Their ages ranged from 20 to 24 years (18 female, mean age = 21.9 years, SD = 1.1). All subjects were right handed and had normal or corrected-to-normal eyesight. They gave written informed consent before the experiment. After the experiment, each participant was paid RMB 15 yuan.

### **2.1.2 Materials**

*The simultaneous same-different task.* The display shown in each trial consisted of a pair of single-digit Arabic numbers ranging from 1 to 9, one on the left and one on the right of the screen. Each possible combination of quantities was presented (except for equal quantities), resulting in  $9 \times 8 = 72$  trials per block. In addition to 72 “different” trials, nine “same” trials were also included which were presented 4 times per block, leading to 36 same and 72 different trials. Three blocks of these 108 experimental trials were presented. The digits were presented in Arial font (13 × 20 mm) using a 19-inch color screen connected to a computer. The left number was presented 4.59 cm to the left of the center of the screen, while the right number was presented 4.59 cm to the right of the center of the screen.

*The sequential same-different task.* The materials were the same as those in the simultaneous same-different task except that the two digits in each trial were presented sequentially.

*Subjective visual similarity questionnaire.* Participants were given a questionnaire in which they rated the subjective visual similarity of all possible 72 pairs of single-

digit Arabic numbers. The Arabic numbers were in Arial size 20 font. Participants were also provided with an instruction sheet to help them understand the 10-point rating scale.

### ***2.1.3 Procedure***

Participants completed all tasks individually in a sound attenuated small room while facing a computer screen from a distance of approximately 60 cm. There were two sessions in the experiment. All participants finished the computer-based task first, and then the questionnaire.

The two variations of the same-different task were computer-based, controlled by E-prime 1.1. All participants finished both simultaneous and sequential same-different tasks. The order of presentation mode, whether simultaneous or sequential, was balanced. Some participants completed the simultaneous same-different task first while others were first to complete the sequential same-different task..

In the simultaneous same-different task, a fixation cross was first presented for 600 ms and then followed by two numbers. The numbers remained on the screen until a button was pressed. The intertrial interval was 800 ms. The subjects were asked to decide as quickly and as accurately as possible whether the two numbers simultaneously presented were the same or different. They were to press the “F” key if the same and the “J” key if different. All the stimuli were randomly presented. Prior to each task, verbal instructions were presented on the screen and there were 16 practice trials before the actual trials. An accuracy feedback was given to subjects. If their accuracy was more than 80%, subjects entered the formal trials. In each task,

after each block, the subjects were allowed to take a break, which they could end at their own pace. This task took approximately 15 min. In the sequential same-different task, the first number appeared for 800 ms after a fixation cross was presented for 600 ms. After a blank screen for 200 ms, the second number appeared on the screen until a button was pressed. Both numbers were presented centrally. The intertrial interval was 800 ms.

Finally, participants were presented with the subjective visual similarity questionnaire. They were instructed to judge the visual similarity of pairs of Arabic numbers using a 10-point scale, from 1 (extremely different) to 10 (totally identical). Participants were told that these judgments were subjective and there were no fixed right or wrong answers. An example of a rating was offered so that participants could understand the instructions well.

#### **2.1.4 Data analyses**

For the same-different task, only the 216 “different” trials were analyzed. For reaction time analyses, incorrect trials were removed (2.2% for the simultaneous presentation and 1.6% for the sequential presentation). Then mean RTs for correct responses were computed, excluding those responses that deviated three or more standard deviations from the mean (1.8% for the simultaneous or sequential presentation).

Two perceptual functions and two numerical functions were employed and calculated:  $P_C$ ,  $P_S$ ,  $N_W$ ,  $N_D$ . The first perceptual function was  $P_C$ . This is Cohen’s (2009) original function,  $P_C = O/D$ , where  $O$  is the number of lines that the two digits share

(i.e. over lapping lines) and  $D$  is the number of difference lines between these two numbers.  $P_S$  is a subjective measure in which each participant rated the perceived visual similarity of every pair of numbers.  $N_W$  is the Welford function used by Cohen (2009) and Garcia-Orza et al. (2012) ( $N_W = \lg (L/L-S)$ , where  $S$  is the smaller number and  $L$  is the larger number).  $N_D$  was based on the distance effect, and the formula was  $N_D = L-S$ .

To find out which function(s) would predict reaction time or accuracy, we conducted simple regressions for each task. Each linear regression had one dependent variable (reaction time or accuracy) and one predictor variable ( $P_C$ ,  $P_S$ ,  $N_W$ , or  $N_D$ ). If the reaction time or accuracy could be predicted by more than one predictor, simultaneous multiple linear regression was used to test the relative contribution of each predictor.

## **2.2 Results**

For the simultaneous same-different task, as seen in Table 1,  $P_S$  predicted both the reaction time and accuracy data. In order to determine the importance of individual predictors, simultaneous multiple linear regression on the reaction time was conducted with two predictor variables ( $P_C$  and  $P_S$ ). The overall model was significant,  $F(2, 69) = 37.981, p < .001, R^2 = .524$ .  $P_S$  was still a significant individual predictor ( $\beta = 0.612, p < .001$ ). However,  $P_C$  was no longer a significant predictor ( $\beta = .179, p = .073$ ). In addition, the variance inflation factor (VIF), a measure of potentially remaining multi-collinearity problems (Cohen, 2003) was smaller than 5 for each predictor (VIFs  $< 1.405$ ). Typically, VIF values larger than 5 or 10 are considered to



indicate multi-collinearity problems. Therefore, no multi-collinearity problem existed in this regression. Fig. 1 provides a plot of the average reaction time and predicted reaction time based on  $P_C$ ,  $P_S$ ,  $N_W$ ,  $N_D$  in the simultaneous same-different task.

---

Tables 1 about here

---

Based on the above results, we hypothesized that the relationship between  $P_C$  and the reaction time may be mediated by  $P_S$ . To further understand the mechanisms underlying relationships among  $P_S$ ,  $P_C$  and the reaction time, we used the INDIRECT procedure for SPSS (see Preacher & Hayes, 2008, for more details) to test our mediation hypothesis. As shown in Fig. 2, the mediation model was estimated to derive the total, direct and indirect effects of  $P_C$  on the reaction time through  $P_S$ . We estimated the indirect effect of  $P_C$  on the reaction time quantified as the product of the OLS regression coefficient estimating  $P_S$  from  $P_C$  (i.e., path  $a$  in Fig. 2), and the OLS regression coefficient estimating the reaction time from  $P_S$ , controlling for  $P_C$  (i.e., path  $b$  in Fig. 2). A bias-corrected bootstrap-confidence interval (CI) for the product of these paths that does not include zero provides evidence of a significant indirect effect (Preacher & Hayes, 2008). Using the INDIRECT procedure with 5000 bootstrap samples revealed a significant positive indirect effect of  $P_C$  on the reaction time through  $P_S$  (point estimate=0.328,  $p < .001$ , 95% percentile, CI = 0.183 to 0.529).

---

Figures 2, 3 about here

-----

For the sequential same-different task, as seen in Table 2, only  $P_S$  predicted the reaction time. Fig. 3 provides a plot of the average reaction time and predicted reaction time based on  $P_C$ ,  $P_S$ ,  $N_W$ ,  $N_D$  in the sequential same-different task.

Furthermore, we compared the regression coefficients of  $P_S$  on the reaction time under simultaneous and sequential presentations. To do this analysis, we first made a dummy variable called “presentation” that was coded 1 for simultaneous presentation and 0 for sequential presentation. In addition, a variable “pres $P_S$ ” was computed as the product of presentation and  $P_S$ . Finally, we used  $P_S$ , presentation and pres $P_S$  as predictors in the regression equation. If the variable pres $P_S$  can significantly predict the reaction time, it indicates that the regression coefficient for sequential presentation is significantly different from that for simultaneous presentation. Results showed that the reaction time was significantly predicted ( $R^2 = .75$ ). All the three predictors had significant prediction on the reaction time ( $ps < .05$ ). More importantly, the prediction of “pres $P_S$ ” on the reaction time was significant ( $\beta = 0.513$ ,  $t = 4.094$ ,  $p < .001$ ), indicating that the regression coefficient for simultaneous presentation was significantly larger than that for sequential presentation.

### **2.3. Discussion**

As we hypothesized, the subjective physical similarity was the best predictor of participants’ performance in the same-different task, regardless of simultaneous or sequential presentation. Although in the simultaneous presentation the objective physical similarity could directly predict the reaction time data, it turned out that the

prediction was mainly mediated by the subjective physical similarity. This finding is consistent with two previous studies (Defever et al, 2012; Wong & Szűcs, 2013), revealing the importance of subjective visual similarity in the symbolic same-different task. This is the first time the subjective physical similarity has been demonstrated to be the primary controlling variable in the sequential same-different task.

In addition, this experiment for the first time showed that the regression coefficient for simultaneous presentation was significantly larger than that for sequential presentation. This finding supports our hypothesis that the contribution of subjective physical similarity was larger in a simultaneous presentation than in a sequential presentation. We discuss the findings of Experiment 1 in more detail in the General Discussion. In Experiment 2, we conduct a traditional magnitude comparison task to further explore the role of subjective physical similarity in a numerical comparison task.

### ***3 Experiment 2: numerical comparison task***

Experiment 2 was designed to expand on the findings of the same-different task in Experiment 1. Here, we conducted a numerical comparison experiment using the same parameters as the simultaneous same-different task in Experiment 1. Cohen et al. (2013) found that the numerical distance was the primary controlling factor in American participants' numerical comparison. However, the objective physical similarity was also revealed to be a significant predictor of participants' comparison performance. This experiment would test their results with Chinese participants.

#### ***3.1 Methods***

### ***3.1.1 Participants***

Forty undergraduates were recruited from Southwest University, China. Their ages ranged from 18 to 24 years (mean age = 21.8 years,  $SD = 1.6$ ). All subjects were of right handedness and had normal or corrected to normal eyesight. They gave written informed consent before the experiment. After the experiment, each participant was paid RMB 15 yuan. These participants did not participate in Experiment 1.

### ***3.1.2 Materials***

The stimuli in the numerical comparison task were the same as those in the “different” trials of the same-different task. There were three blocks with each block 72 trials.

### ***2.1.3 Procedure***

The procedure was highly similar as that in Experiment 1. Only the instructions differed. The subjects had to decide which number was numerically bigger and consequently press the corresponding button (“F” if the left number was larger and “J” if the right number was larger). This task also took approximately 15 min. All participants completed the subjective visual similarity questionnaire.

## ***2.2 Results***

Similarly to Experiment 1, three participants were excluded because their responses in the questionnaire were systematically biased (all responses were the same). For RT analyses, incorrect trials were removed (2.6%). The mean RTs for correct responses were computed, excluding those responses that deviated three or more standard

deviations from the mean for each task (1.8%).

As seen in Table 3, two functions including  $N_W$  and  $N_D$  significantly predicted the reaction time and accuracy. In addition,  $P_S$  predicted the reaction time data. Similarly, in order to determine the importance of individual predictors, simultaneous multiple linear regressions were used with three predictors ( $N_W$ ,  $N_D$  and  $P_S$ ). For the reaction time data, the overall model was significant,  $F(3, 68) = 60.391, p < .001, R^2 = .727$ . Only  $N_W$  was significant as an individual predictor ( $\beta = 0.712, p < .001$ ).  $N_D$  and  $P_S$  showed negligible  $\beta$  values ( $\beta \leq .03, p \geq .128$ ). For the accuracy data, the overall model was significant,  $F(2, 69) = 40.472, p < .001, R^2 = .540$ . Only  $N_W$  was a significant individual predictor ( $\beta = -.761, p < .001$ ).  $N_D$  and  $P_S$  showed negligible  $\beta$  values ( $\beta = -.034, p = .796$ ). In addition, in both reaction times and accuracy analyses, the VIF was smaller than 5 for each predictor (VIFs < 2.836), indicating no multicollinearity problem. Fig. 4 provides a plot of the average reaction time and predicted reaction time based on  $P_C$ ,  $P_S$ ,  $N_W$ ,  $N_D$  in the numerical comparison task.

-----  
Table 2 about here  
-----

These results allow us to propose the possible mediator effect of  $N_W$  between  $P_S$  and the reaction time (Fig. 5). To further understand the mechanisms underlying relationships among  $P_S$ ,  $N_W$  and the reaction time, we used the INDIRECT procedure within SPSS to test our mediation hypothesis. Using the INDIRECT procedure with 5000 bootstrap samples revealed a significant positive indirect effect of  $P_S$  on the

reaction time through  $N_W$  (point estimate=0.285,  $p = .003$ , 95% percentile, CI=0.099 to 0.471). In addition, as seen in Fig. 5, the direct effect of  $P_S$  on the reaction time was not significant ( $p = .745$ ).

-----  
Figures 4, 5 about here  
-----

### **3.3. Discussion**

This experiment showed that  $N_W$  was the best predictor for numerical comparison ability, similarly to the study by Cohen et al. (2013). However, our results are different from those in two ways. First, it is subjective, not objective, physical similarity which predicts the reaction time in the numerical comparison task. Second, we analyzed the possible mediator effect of numerical distance in the prediction of subjective physical similarity, which is not done in the study by Cohen et al. (2013). Results showed that the influence of subjective physical similarity was mediated by numerical distance. This finding indicates that neither objective nor subjective physical similarity was activated in the numerical comparison task.

In Experiment 3, we expand our findings to the physical comparison task. The task explicitly required participants to compare physical magnitudes. We aimed to explore the contribution of numerical distance and physical similarity in participants' physical comparison.

### **4 Experiment 3: physical comparison task**

Experiment 3 was designed to expand on the findings of Experiments 1 and 2

using the physical comparison task. This is the first study to explore the influence of numerical distance and physical similarity on physical magnitude comparison.

## ***4.1 Methods***

### ***4.1.1 Participants***

Twenty undergraduates were recruited from Southwest University, China. Their ages ranged from 18 to 25 years (mean age = 21.5 years, SD = 1.5). All subjects were of right handedness and had normal or corrected-to-normal eyesight. They gave written informed consent before the experiment. After the experiment, each participant was paid RMB 15 yuan. These participants did not participate in Experiments 1 and 2.

### ***4.1.2 Materials***

The numbers presented were in one of two physical sizes: large (15 × 23 mm) and small (13 × 20 mm). There were six blocks with per block 72 trials. In each block, half of the trials were in the congruent condition (with the numerically larger number appearing in a physically larger size) while the other half were in the incongruent condition (with the numerically larger number appearing in a physically smaller size). In each block, the physically larger numbers appeared on the left side in half trials and appeared on the right side in half trials.

### ***4.1.3 Procedure***

The procedure was similar to that in Experiment 2 except for the instructions. In the physical comparison task, the subjects were asked to decide as quickly and as accurately as possible which number was physically bigger (“F” if the left number

physically larger and “J” if the right number physically larger). This task also took approximately 15 min. Then all participants finished the subjective visual similarity questionnaire.

## 4.2 Results

Similarly to Experiment 1, four participants were excluded because of their low accuracy in the incongruent condition (less than 10%). For the reaction time analyses, incorrect trials were removed (9.2%). The mean RTs for correct responses were computed, excluding those responses that deviated three or more standard deviations from the mean for each condition (1.9%).

As seen in Table 4, no function predicted the reaction time data. Three functions including  $P_S$ ,  $N_W$  and  $N_D$  significantly predicted accuracy. Simultaneous multiple linear regression on accuracy with three predictor variables ( $P_S$ ,  $N_W$  and  $N_D$ ) showed that  $P_S$  was still a significant predictor ( $\beta = 0.267$ ,  $p = .029$ ). In contrast,  $N_D$  or  $N_W$  could no longer predict the accuracy ( $ps > .128$ ). In addition, the VIF was smaller than 5 for each predictor (VIFs  $< 2.798$ ), indicating no multi-collinearity problem in this regression. However, when simultaneous multiple linear regression on accuracy with two predictor variables ( $P_S$  and  $N_D$ ), both  $P_S$  and  $N_D$  were significant predictors ( $\beta = .267$ ,  $p = .023$ ;  $\beta = -.272$ ,  $p = .021$ ). In contrast, when simultaneous multiple linear regression on accuracy with two predictor variables ( $P_S$  and  $N_W$ ),  $P_S$  was a significant predictor ( $\beta = .272$ ,  $p = .027$ ) but  $N_W$  was not ( $\beta = .212$ ,  $p = .084$ ). Fig. 6 provides a plot of the average reaction time and predicted reaction time based on  $P_C$ ,  $P_S$ ,  $N_W$ ,  $N_D$  in the physical comparison task.



-----  
Table 3 about here  
-----

The RT and accuracy was also subjected to ANOVAs with congruity as the between-subject factor. The reaction time results indicated that the main effect of congruity was significant,  $F(1, 15) = 95.450$ ,  $\eta^2 = .864$ , Cohen's  $d = 5.828$ ,  $p < .001$ , with the incongruent condition (499 ms) slower than the congruent condition (441 ms). The accuracy results additionally revealed a significant congruity main effect,  $F(1, 15) = 36.019$ ,  $\eta^2 = .706$ , Cohen's  $d = 1.983$ ,  $p = .002$ , with greater accuracy (99%) in the congruent condition than in the incongruent condition (94%). All these results showed that the numerical magnitude was automatically represented in the physical comparison task.

#### **4.3. Discussion**

This experiment showed that  $P_S$  was a predictor of participants' physical comparison abilities. However, data also showed that  $N_D$  was another predictor of participants' physical comparison. This is based on the finding that  $N_D$  was a significant predictor not only in simple regression but also in simultaneous multiple regression with two predictor variables ( $P_S$  and  $N_D$ ). Consistently, our data revealed that the numerical magnitude was automatically represented in the physical comparison task since participants performed better in the congruent condition than in the incongruent condition. Taken together, both subjective physical similarity and numerical magnitude contributed to the physical comparison task. In Experiment 4,

we expand our findings to the priming task.

### ***5 Experiment 4: priming task***

Experiment 4 was designed to expand on the findings of Experiments 1, 2, and 3 using the priming task. This is also the first study to explore the influence of numerical distance and physical similarity with the priming task.

#### ***5.1 Methods***

##### ***5.1.1 Participants***

Twenty undergraduates were recruited from Southwest University, China. Their ages ranged from 19 to 26 years (mean age = 21.3 years, SD = 1.7). All subjects were of right handedness and had normal or corrected to normal eyesight. They gave written informed consent before the experiment. After the experiment, each participant was paid RMB 15 yuan. These participants did not participate in Experiments 1, 2 or 3.

##### ***5.1.2 Materials***

In this task, the prime and target stimuli consisted of Arabic numbers 1-9, excluding 5. As a result, there were  $8 \text{ prime} \times 7 \text{ target} = 56$  possible prime-target combinations from distance 1 (e.g. prime 2- target 1) to distance 3 (e.g., prime 6 - target 9). Among these trials, 24 were of the congruent type (prime and target both falling on the same side of 5) and 32 trials were incongruent. Six blocks of experimental trials were presented containing 24 congruent and 32 incongruent trials. In congruent trials, the prime and target evoked the same response (for instance, prime 2 followed by target 4: both smaller than 5) while in incongruent trials, the

prime and target evoked different responses (e.g., prime 2- target 6; the first is “smaller than five”, the second “larger than five”).

### **5.1.3 Procedure**

In the priming task, each trial began with a forward mask which appeared in the middle of screen for 57 ms. This mask consisted of six hash signs (#) of the same size and font as the Arabic targets. Then the prime was presented for 57 ms, followed by a backward mask for another 57 ms. The backward mask was the same as the forward mask. Finally, the target was presented and remained on the screen until the participant responded. The font of the primes (8 mm wide × 10 mm high) was smaller than that of the targets (10 mm wide × 15 mm high). By doing so, it could reduce the physical overlap between prime and target (Reynvoet, Brysbaert, & Fias, 2002). All stimuli were presented in black on a white background and were centered on the screen. All stimuli were presented in Arial Unicode MS font. Participants were asked to decide as quickly and as accurately as possible whether the number on the screen was smaller or larger than five by pressing the “F” key if the number was smaller than five and the “J” key if the number was larger than five. This task also took approximately 15 min.

### **5.2 Results**

One participant was excluded for his low accuracy (less than 80%). For reaction time analyses, incorrect trials were removed (3.4%). In addition, scores more than three standard deviations from the mean were excluded as outliers (1.8 %).

-----

Table 4 about here

-----

As seen in Table 4,  $N_W$  and  $N_D$  significantly predicted the reaction time demonstrating a classical priming distance effect with better performance in close distance pairs than in far distance pairs. Simultaneous multiple linear regression on accuracy with two predictor variables ( $N_W$  and  $N_D$ ) showed that the overall model was significant,  $F(2, 53) = 3.705$ ,  $p = .031$ ,  $R^2 = .123$ . The VIF (2.520) was smaller than 5 for each predictor, indicating no multi-collinearity problem in this regression. However,  $N_D$  was no longer a significant predictor,  $\beta = -.005$ ,  $p = .980$ , and the prediction of  $N_W$  was also not significant,  $\beta = -.354$ ,  $p = .089$ . These results suggested that both predictors accounted for the same variance (i.e., redundancy), making it impossible to assess which was the important factor (see Field, 2005, p. 175; Meert et al., 2009). Fig. 7 provides a plot of the average reaction time and predicted reaction time based on  $P_C$ ,  $P_S$ ,  $N_W$ ,  $N_D$  in the priming task.

In addition, in order to confirm whether the classic priming effect appeared, the reaction time and accuracy was subjected to ANOVAs with congruity as the within-subject factor. Results showed a significant behavioral priming effect,  $F(1, 18) = 44.066$ ,  $\eta^2 = .710$ , Cohen's  $d = 0.453$ ,  $p < .001$ , the incongruent condition (495ms) was slower than the congruent condition (474 ms). The accuracy results did not reveal a significant congruity main effect ( $p = .328$ ).

### 5.3. Discussion

This experiment showed that both  $N_W$  and  $N_D$  were controlling factors when

participants performed the priming task. Neither objective nor subjective physical similarity had influence on participants' performance in the priming task.

## ***6 General Discussion***

The data from Experiments 1–4 provides some insight into how Arabic numbers are processed in different tasks. Specifically, participants make their judgments based on perceptual or semantic magnitude information. If physical similarity contributes to participants' performance in a task, it suggests that participants carry out the task based on perceptual information. If numerical distance contributes to participants' performance in a task, it suggests participants carry out the task based on semantic magnitude information. Based on such logical reasoning, this study has four important findings. First, perceptual information was involved in the same-different task; however, it is the subjective, but not objective, physical similarity that contributed to the same-different judgment. Second, whether perceptual information was involved depends on task demands. In the same-different task, only perceptual information was involved. In the physical comparison task, both perceptual and semantic magnitude information was involved since both subjective physical similarity and numerical distance contributed to participants' performance. However, in numerical comparison and priming tasks, only semantic magnitude information was involved since only numerical distance contributed to participants' performance. Third, this study observed that the presentation mode in the same-different task had influence on the contribution of physical similarity in the processing of single-Arabic numbers. Specifically, the contribution of subjective physical similarity was larger in a

simultaneous presentation than in a sequential presentation. Fourth, the semantic magnitude representation of symbolic numbers depended on task demands.

Participants tended to use log representations in numerical comparison and linear representations in the physical comparison task. In the priming task, participants did not show an obvious tendency to mainly rely on log or linear representations. Below we discuss underlying reasons for these results and important implications of these results.

This study indicates that subjective, but not objective, physical similarity contributes to same-different judgments. It seems that the objective visual similarity proposed by Cohen (2009) does not best describe the perceptual structures of digits. As revealed by mediator analyses in our study, the contribution of objective visual similarity in the same-different task found by Cohen (2009) might be mediated by subjective physical similarity.

What could be driving subjective visual similarity to influence in this way? It is likely that participants made judgments based on vague and gist perceptions rather than accurate and verbatim perceptions (Reyna, & Brainerd, 1995). For example, curves and straight lines that form acute angles might drive subjective visual similarity (Wong, & Szűcs, 2013). Indeed, we verified this assumption in a preliminary way in our study. We defined those number pairs with curves as the having the highest similarity, including 30 pairs constructed by 2, 3, 5, 6, 8, and 9. Those number pairs with straight lines were defined as the moderate similarity, including 12 pairs constructed by 1, 4, 7, and 9 as well as 8 pairs consisted of 2-4, 4-2,

2-7, 7-2, 5-7, 7-5, 4-5, 5-4. Both numbers in each pair have vertical lines or transverse lines. The left 22 pairs were defined as the lowest similarity. A simple regression analysis on the subjective physical similarity with the defined similarity as a predictor showed that our defined similarity could explain 41% -54% variance of the subjective physical similarity. In Experiment 1,  $R^2 = .502$ ,  $\beta = .713$ ,  $p < .001$ ; in Experiment 2,  $R^2 = .537$ ,  $\beta = .737$ ,  $p < .001$ ; in Experiments 3 and 4,  $R^2 = .410$ ,  $\beta = 0.647$ ,  $p < .001$ . In contrast, the objective physical similarity defined by Cohen (2009) could only explain 28.3% variance of the subjective physical similarity,  $R^2 = .283$ ,  $\beta = 0.537$ ,  $p < .001$ . Future research is needed to more accurately describe the source of subjective visual similarity.

The second important finding is that the contribution of physical similarity depends on task demands. One possible explanation framework for this finding is Dehaene's (1992) Triple Code Model. Dehaene (1992) suggested that three mental representations of number: an auditory verbal representation, a visual Arabic representation, and an analog magnitude representation. In the course of number processing, numbers may be transcoded from one internal representation to the other. In numerical comparison, physical comparison and priming tasks, Arabic numbers are internally transcoded to the analog magnitude representation; however, in the same-different task, such transcoding does not occur.

Specifically, the transcoding from visual Arabic numbers to analog magnitude occurs since all the three tasks including numerical comparison, physical comparison and priming tasks explicitly require a quantitative comparison process (bigger or

smaller). Numerical comparison and priming tasks explicitly require a comparison process of numerical magnitude whereas the physical comparison task explicitly entails a comparison process of physical size. It is likely that as long as a comparison process is required, regardless of numerical or physical, the numerical magnitude will be represented intentionally or automatically. The comparison process entails constructing a continuous mental number line and then retrieving analog values for each compared digit (Moyer & Landauer, 1967). However, the same-different task only entails a qualitative identification and classification process (is or not). In the task, it is unnecessary for participants to construct a continuous mental number line and retrieve analog values for each digit. The participants simply need to classify the surface structures of the presented digits into “same” or the “different” categories. Therefore, the transcoding from visual Arabic numbers to analogical magnitude is not necessary and does not occur.

It should be noted that in addition to numerical distance, subjective physical similarity also contributed to participants’ physical comparison. This finding is easy to understand since this task explicitly requires participants to compare physical size. If a physical same-different task is administered to participants, it is likely that the subjective physical similarity will contribute to participants’ performance. However, since the comparison process is not entailed in such a task, the numerical distance will not contribute to participants’ comparison. Indeed, using the physical same-different task, Ganor-Stern and Tzelgov (2008) did not observe the numerical distance effect, regardless of that the numbers were presented in the Arabic or Indian or mixed



notation. Future research is needed to test whether the subjective physical similarity will contribute to participants' performance in the physical same-different task.

The third important finding is that the contribution of subjective physical similarity was larger in a simultaneous presentation than in a sequential presentation. This finding is consistent with our hypothesis. One possible reason is the relative speed account of physical size and number processing (MacLeod, 1991; Schwarz & Ischebeck, 2003). Specifically, in the simultaneous same-different task, the time course of semantic information is slower than that of perceptual information. Therefore, participants give their responses on the basis of perceptual similarity rather than semantics for the simultaneous presentation. In contrast, the sequential representation would afford a greater opportunity to accessing semantic information and hence to rely less on visual information since one number is not presented visually at the comparison stage. Such an explanation is consistent with the study by Frick (1985) in which it was found that more visual letter confusion errors occurred for a simultaneous presentation but not for a sequential presentation. The author concluded that the sequential presentation is not very successful in eliciting use of a visual short-term store as compared with the simultaneous presentation.

Finally, this study showed that whether symbolic numbers were represented on a linear scale or a log scale depends on task demands. Log representations were mainly relied on in the numerical comparison task and linear representations in the physical comparison task. The possible reasons for these findings are from two aspects. On the one hand, as compared to linear representations, log representations can maintain

ordinal properties of numbers (Laski & Siegler, 2007). On a log scale, estimates exaggerate differences in the magnitudes of smaller numbers and compress differences in the magnitudes of larger ones. In contrast, on a linear scale, estimates neither exaggerate nor compress differences among numbers (Siegler & Booth, 2004). Therefore, logarithmic representations are ideal for representing situations in which differences at the low end of the range are paramount, whereas linear representations are ideal when all parts of the range are equally important (Siegler & Opfer, 2003). The order of numbers is such an example that logarithmic representations should be suitable since those numbers which temporally come earlier in the verbal counting series are more frequently used and paramount.

On the other hand, it has been suggested that not only the magnitude information but also the numerical order information is involved in the numerical comparison task (Turconi, Campbell, & Seron, 2006). The numerical distance effect was found to be smaller for ascending (e.g. 2 vs. 5) compared to descending pairs (e.g. 5 vs. 2). In this study, we compared the performance for ascending and descending pairs in the numerical comparison task. The results showed that the reaction time for ascending pairs (425 ms) was faster than for descending pairs (438 ms),  $F = 4.410$ ,  $p = .039$ , suggesting the involvement of direction-related order information in magnitude comparison. As such, people tend to use logarithmic representations on numerical magnitude comparison tasks. In the physical comparison task, it is likely that only magnitude information of numbers is unintentionally processed based on the numerical magnitude of numbers being irrelevant to the task requirement. Therefore,

linear representations were mainly relied on in the physical comparison task.

Different from numerical and physical comparison tasks, the priming task produced the reversed distance effect ( $N_D$  positively predicted the reaction time). A reverse distance effect has been typically observed in order judgment tasks which require participants to judge whether two or three numbers are in the correct order (Franklin & Jonides, 2009; Turconi, Campbell, & Seron, 2006). Therefore, as done in the order task, it is likely that a serial scanning occurred in the priming task (faster response for adjacent than non-adjacent pairs in the sequence) leading to a reverse distance effect (Franklin & Jonides, 2009). For the finding that participants did not show an obvious tendency to primarily rely on log or linear representations in the priming task, it may be due to the statistical limit of the high correlation between  $N_D$  and  $N_W$  leading to the redundancy effect. However, when we compared participant's performance between ascending and descending pairs, no significant differences were found in the priming task ( $ps > .426$ ) suggesting that participants possibly did not process the direction-related order information. It should be noted that our results were limited to the current priming paradigm.

The results of the present investigation have important implications for researchers who use these tasks to study individual and group differences. Specifically, using an unreliable measure makes it more difficult to detect group differences in numerical representations. For example, the same-different task, regardless of sequential or simultaneous presentation, is not an adequate measure to study the development of magnitude representations since the judgments are based on the

physical similarity of digits. However, such a conclusion is limited to the case when both stimuli are Arabic numbers. It has been found that when the same-different task involves mixed notations (e.g., Arabic and verbal number notations; symbolic Arabic and non-symbolic arrays of dots), the numerical representation can be accessed (Defever, Sasanguie, Vandewaetere, Reynvoet, 2012; Van Opstal & Verguts, 2011).

The other three tasks excluding the same-different task (numerical comparison, physical comparison and priming tasks) do seem to elicit the numerical distance effect. It has been disputed whether the numerical comparison task is a reliable metric of numerical representation. Van Opstal et al. (2008) demonstrated that while a comparison distance effect can be found for both letters and numbers, a priming distance effect is only observed for numbers but not for letters. They concluded that the comparison distance effect is due to a comparison process being involved but not due to the numerical representation being accessed. However, this conclusion should be considered cautiously as it is likely that order processing drives the comparison distance effect for numbers and letters, since both numbers and letters entail order information. Such a possibility is not considered in their study.

## References

- Buckley, P. B., Gillman, C. B. (1974). Comparisons of digits and dot patterns. *Journal of Experimental Psychology*, 103, 1131–1136.
- Campbell, J. I. D., & Clark, J. M. (1988). An encoding-complex view of cognitive number processing: Comment on McCloskey, Sokol, and Goodman (1986). *Journal of Experimental Psychology, General*, 117, 204–214.

- Cohen, D. (2003). Applied multiple regression/correlation analysis for the behavioral sciences (3rd edn.). London: Routledge.
- Cohen, D. J. (2009). Integers do not automatically activate their quantity representation. *Psychonomic Bulletin and Review*, 16, 332–336.
- Cohen, D. J., Warren, E., & Blanc-Goldhammer, D. (2013). Cross-format physical similarity effects and their implications for the numerical cognition architecture. *Cognitive Psychology*, 66, 355-379.
- Cohen Kadosh, R., Henik, A., & Rubinsten, O. (2008). Are Arabic and verbal numbers processed in different ways? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 1377-1391.
- Defever, E., Sasanguie, D., Vandewaetere, M., & Reynvoet, B. (2012). What can the same–different task tell us about the development of magnitude representations? *Acta Psychologica*, 140, 35-42.
- Dehaene, S. (1989). The psychophysics of numerical comparison: A reexamination of apparently incompatible data. *Perception & Psychophysics*, 45, 557-566.
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, 44, 1-42.
- Dehaene, S., & Akhavein, R. (1995). Attention, automaticity and levels of representation in number processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 314–326.
- Domahs, F., Klein, E., Moeller, K., Nuerk, H. -C., Yoon, B. -C., & Willmes, K. (2012). Multimodal semantic quantity representations: Further evidence from Korean sign language. *Frontiers in Psychology*, 2, 1-10.

- Duncan, E.M., &McFarland, C. E. (1980). Isolating the effects of symbolic distance and semantic congruity in comparative judgments: An additive-factors analysis. *Memory & Cognition*, 8, 612-622.
- Field, A. (2005). *Discovering statistics using SPSS*. London: Sage.
- Franklin, M. S., & Jonides, J. (2009). Order and magnitude share a common representation in parietal cortex. *Journal of Cognitive Neuroscience*, 21(11), 2114-2120.
- Frick, R. W. (1985). Testing visual short-term memory: Simultaneous versus sequential presentations. *Memory & Cognition*, 13 (4), 346-356.
- Henik, A., &Tzelgov, J. (1982). Is three greater than five: The relation between physical and semantic size in comparison tasks. *Memory and Cognition*, 10, 389-395.
- Hinrichs, J. V., Yurko, D. S., Hu, J. (1981). Two-digit number comparison: Use of place information. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 890–901.
- Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children’s mathematics achievement. *Journal of Experimental Child Psychology*, 103, 17-29.
- Ganor-Stern, D. Pinhas, M., & Tzelgov, J. (2009). Comparing two-digit numbers: The importance of being presented together. *Quarterly Journal of Experimental Psychology*, 62(3), 444-452.
- Ganor-Stern, D., & Tzelgov, J. (2008). Across-notation automatic numerical processing.

*Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 430-437.

Garcia-Orza, J., Perea, M., Abu Mallouh, R., & Carreiras, M., (2012). Physical similarity (and not quantity representation) drives perceptual comparison of numbers: Evidence from two Indian notations. *Psychonomic Bulletin and Review*, 19, 294-300.

Kaufmann, L., Koppelstaetter, F., Delazer, M., Siedentopf, C., Rhomberg, P., Golaszewski, S., et al. (2005). Neural correlates of distance and congruity effects in a numerical Stroop task: An event-related fMRI study. *NeuroImage*, 25, 888-898.

Laski, E. V., & Siegler, R. S. (2007). Is 27 a big number? Correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison. *Child Development*, 78, 1723-1743.

Lyons, I. M., & Ansari, D. (2008). The cerebral basis of mapping nonsymbolic numerical quantities onto abstract symbols: An fMRI training study. *Journal of Cognitive Neuroscience*, 21, 1720–1735.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109, 163-203.

Meert, G., Grégoire, J., & Noël, M.-P. (2009). Rational numbers: Componential versus holistic representation of fractions in a magnitude comparison task. *The Quarterly Journal of Experimental Psychology*, 62(8), 1598-1616.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical

- inequality. *Nature*, 215, 1519–1520.
- Moyer, R. S., Landauer, T. K. (1973). Determinants of reaction time for digit inequality judgments. *Bulletin of the Psychonomic Society*, 1, 167–168.
- Notebaert, K., Pesenti, M., & Reynvoet, B. (2010). The neural origin of the quantity priming effect: Distance-dependent recovery of intraparietal activation using symbolic magnitudes. *Human Brain Mapping*, 31, 669–677.
- Perry, J. R. & Lupker, S. J. (2012). An investigation of the time course of category congruence and priming distance effects in number classification tasks. *Canadian Journal of Experimental Psychology*, 66, 193-203.
- Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, 7, 1-75.
- Reynvoet, B., Brysbaert, M., Fias, W. (2002). Semantic priming in number naming. *Quarterly Journal Experimental Psychology: Section A*, 2002, 55(4):1127-1139.
- Reynvoet, B., Caessens, B., & Brysbaert, M. (2002). Automatic stimulus-response associations may be semantically mediated. *Psychonomic Bulletin & Review*, 9, 107-112.
- Reynvoet, B., De Smedt, B., Van den Bussche, E. (2009). Children's representation of symbolic magnitude: the development of the priming distance effect. *Journal of Experimental Child Psychology*, 103, 480-489.
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40, 879-891.



- Schwarz, W., & Ischebeck, A. (2003). On the relative speed account of number-size interference in comparative judgments of numerals. *Journal of Experimental Psychology: Human Perception & Performance*, 29, 507-522.
- Shepard, R. N., Kilpatrick, D.W., Cunningham, J. P. (1975). The internal representation of numbers. *Cognitive Psychology*, 7, 82-138.
- Siegler, R. S., & Booth, J. (2004). Development of numerical estimation in young children. *Child Development*, 75, 428-444.
- Siegler, R. S., & Opfer, J. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science*, 14, 237-243.
- Szűcs, D., & Soltesz, F. (2007). Event-related potentials dissociate facilitation and interference effects in the numerical Stroop paradigm. *Neuropsychologia*, 45, 3190-3202.
- Szűcs, D., & Soltesz, F. (2008). The interaction of task-relevant and task-irrelevant stimulus features in the number/size congruency paradigm: An ERP study. *Brain Research*, 1190, 143-158.
- Turconi, E., Campbell, J. I. D., & Seron, X. (2006). Numerical order and quantity processing in number comparison. *Cognition*, 98(3), 273-285.
- Tzelgov, J., Meyer, J., & Henik, A. (1992). Autonomic and intentional processing of numerical information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 166-179.
- Turconi, E., Campbell, JI. , Seron, X. (2006). Numerical order and quantity processing

- in number comparison. *Cognition*, 98, 273-278
- Van Opstal, F., Gevers, W., De Moor, W., & Verguts, T. (2008). Dissecting the symbolic distance effect: priming and comparison distance effects in numerical and non-numerical orders. *Psychonomic Bulletin & Review*, 15, 419-425.
- Verguts, T., & Fias, W. (2004). Representation of number in animals and humans: A neural model. *Journal of Cognitive Neuroscience*, 16(9), 1493-1504.
- Van Opstal, F., & Verguts, T. (2011). The origins of the numerical distance effect: The same different task. *Journal of Cognitive Psychology*, 23, 112-120.
- Verguts, T., & van Opstal, F. (2005). Dissociation of the distance effect and size effect in one-digit numbers. *Psychonomic Bulletin & Review*, 12, 925-930.
- Welford, A. T. (1960). The measurement of sensory-motor performance: Survey and reappraisal of twelve years' progress. *Ergonomics*, 3, 189-230.
- Wong, B., & Szűcs, D. (2013). Single-digit Arabic numbers do not automatically activate magnitude representations in adults or in children: Evidence from the symbolic same–different task. *Acta Psychologica*, 144, 488-498.
- Young, C. J., & Opfer, J. E. (2011). Psychophysics of numerical representation: A unified approach to single- and multi-digit magnitude estimation. *Journal of Psychology*, 219, 58-63.
- Zhang, J., & Wang, H. (2005). The effect of external representations on numeric tasks. *Quarterly Journal of Experimental Psychology*, 58, 817-838.
- Zhou, X., Chen, C., Chen, L., & Dong, Q. (2008). Holistic or compositional representation of two-digit numbers? Evidence from the distance, magnitude, and

SNARC effects in a number-matching task. *Cognition*, 106, 1525-1536.

**Table 1**

Results of linear regressions for the same-different task in Experiment1.

		The reaction times analyses				Accuracy analyses			
		$R^2$	$\beta$	$F$	$p$	$R^2$	$\beta$	$F$	$p$
Simultaneous Presentation	P <sub>S</sub>	.501	.708	70.319	.000**	.215	-.463	19.126	.000**
	P <sub>C</sub>	.258	.508	24.303	.000**	.041	-.203	3.000	.088
	N <sub>W</sub>	.020	.140	1.399	.241	.024	-.156	1.750	.190
	N <sub>D</sub>	.009	-.095	.636	.428	.001	.038	.102	.751
Sequential Presentation	P <sub>S</sub>	.062	.250	4.647	.035*	.005	-.071	.358	.552
	P <sub>C</sub>	.046	.214	3.353	.071	.001	-.027	.053	.819
	N <sub>W</sub>	.014	.118	0.986	.324	.007	.081	.460	.500
	N <sub>D</sub>	.004	-.060	.256	.614	.015	-.123	1.080	.302

**Table 2**

Results of linear regressions for the numerical comparison task in Experiment 2.

	The reaction times analyses				Accuracy analyses			
	$R^2$	$\beta$	$F$	$p$	$R^2$	$\beta$	$F$	$p$
P <sub>S</sub>	.094	.307	7.280	.009**	.004	-.064	.284	.596
P <sub>C</sub>	.043	.208	3.169	.079	.000	-.019	.026	.872
N <sub>W</sub>	.717	.847	177.453	.000**	.539	-.734	81.969	.000**
N <sub>D</sub>	.526	-.726	77.835	.000**	.319	.565	32.808	.000**

**Table 3**

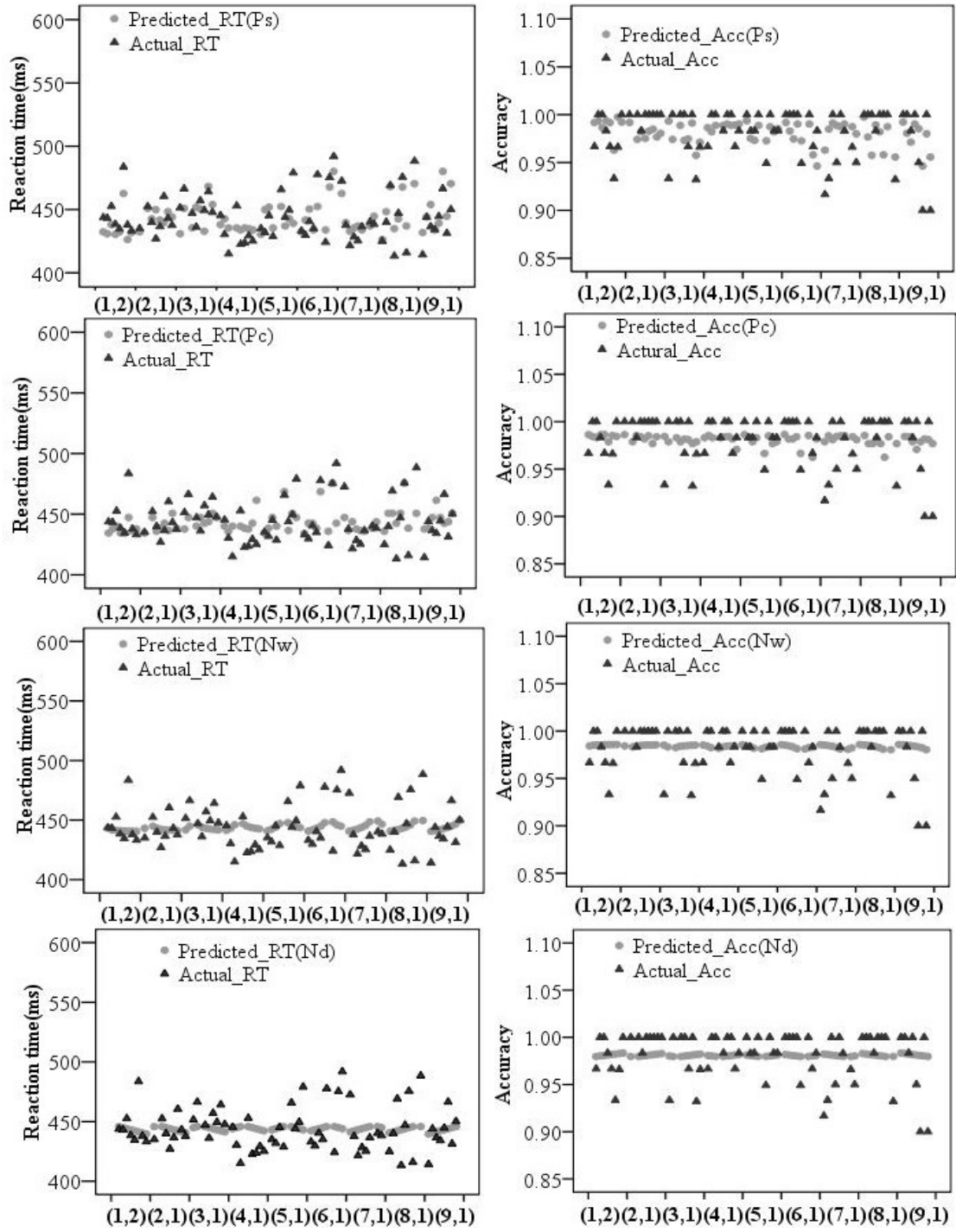
Results of linear regressions for the physical comparison task in Experiment 3.

	The reaction times analyses				Accuracy analyses			
	$R^2$	$\beta$	$F$	$p$	$R^2$	$\beta$	$F$	$p$
P <sub>S</sub>	.017	-.131	1.218	.273	.129	.359	10.366	.002*
P <sub>C</sub>	.001	-.027	.053	.819	.023	.152	1.659	.202
N <sub>W</sub>	.000	.012	.010	.921	.105	.323	8.181	.006*
N <sub>D</sub>	.026	.160	1.836	.180	.131	-.362	10.587	.002**

**Table 4**

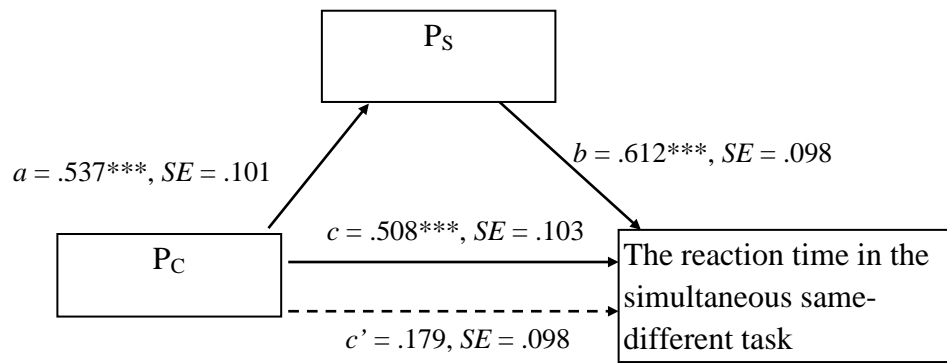
Results of linear regressions for the priming task in Experiment 4.

	The reaction times analyses				Accuracy analyses			
	$R^2$	$\beta$	$F$	$p$	$R^2$	$\beta$	$F$	$p$
P <sub>S</sub>	.061	-.247	3.511	.066	.003	-.054	.157	.693
P <sub>C</sub>	.008	-.090	.437	.512	.009	.094	.483	.490
N <sub>W</sub>	.123	-.350	7.549	.008**	.002	-.047	.121	.729
N <sub>D</sub>	.073	.270	4.245	.044*	.006	.075	.307	.582

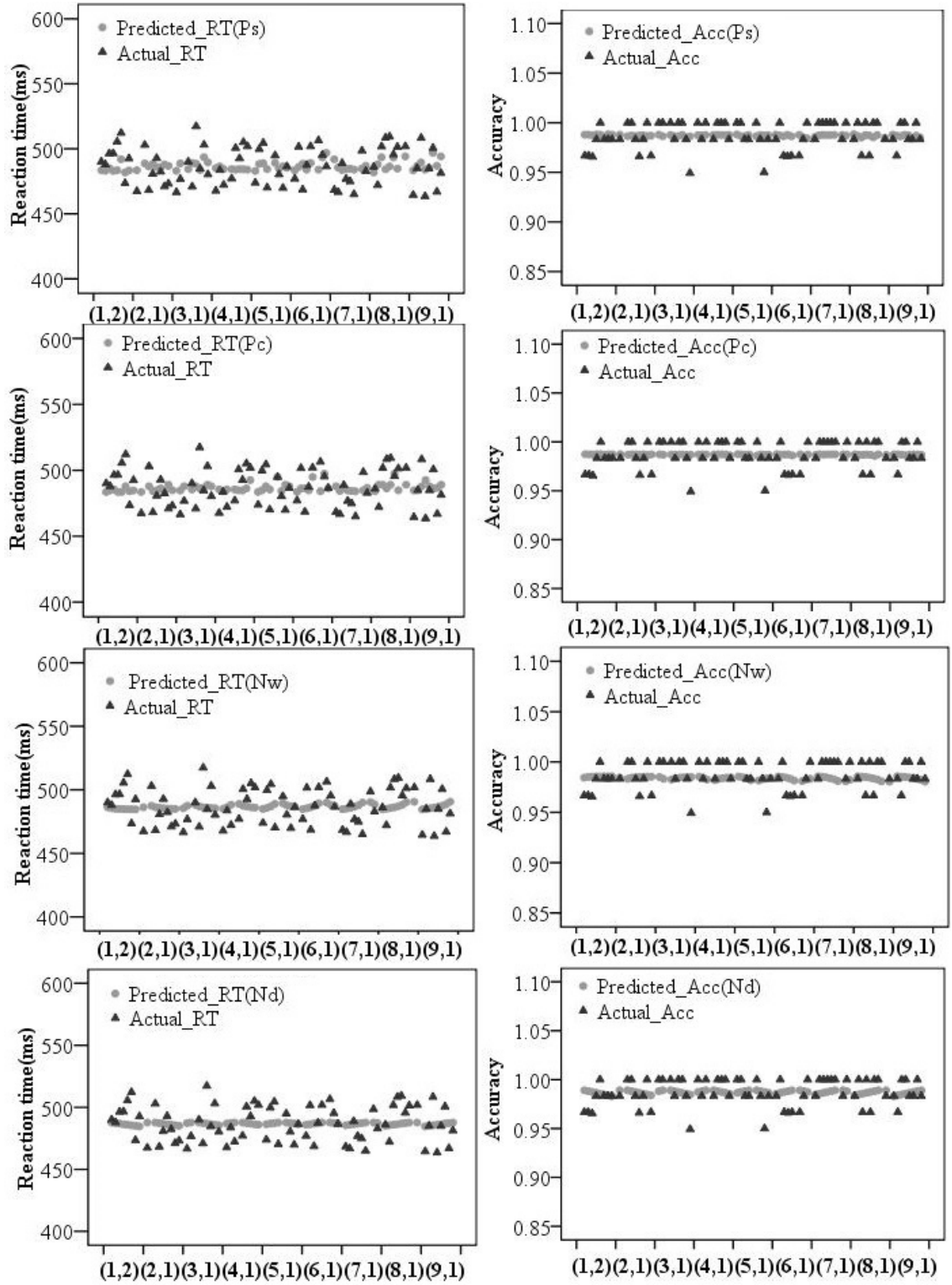


**Fig.1.** Plot of actual average RT/Accuracy and predicted RT/Accuracy based on  $P_s$ ,  $P_c$ ,  $N_w$ ,  $N_d$  for the simultaneous same-different task (Experiment 1). On the x-axis, (x, y) indicates the pair of “different” digits that was presented. For left to right, the pairs progress from (1, 2), (1, 3), (1, 4) ... (1, 9), (2, 1), etc. Only  $P_s$  was a significant predictor of RT/Accuracy in simultaneous multiple linear regression.

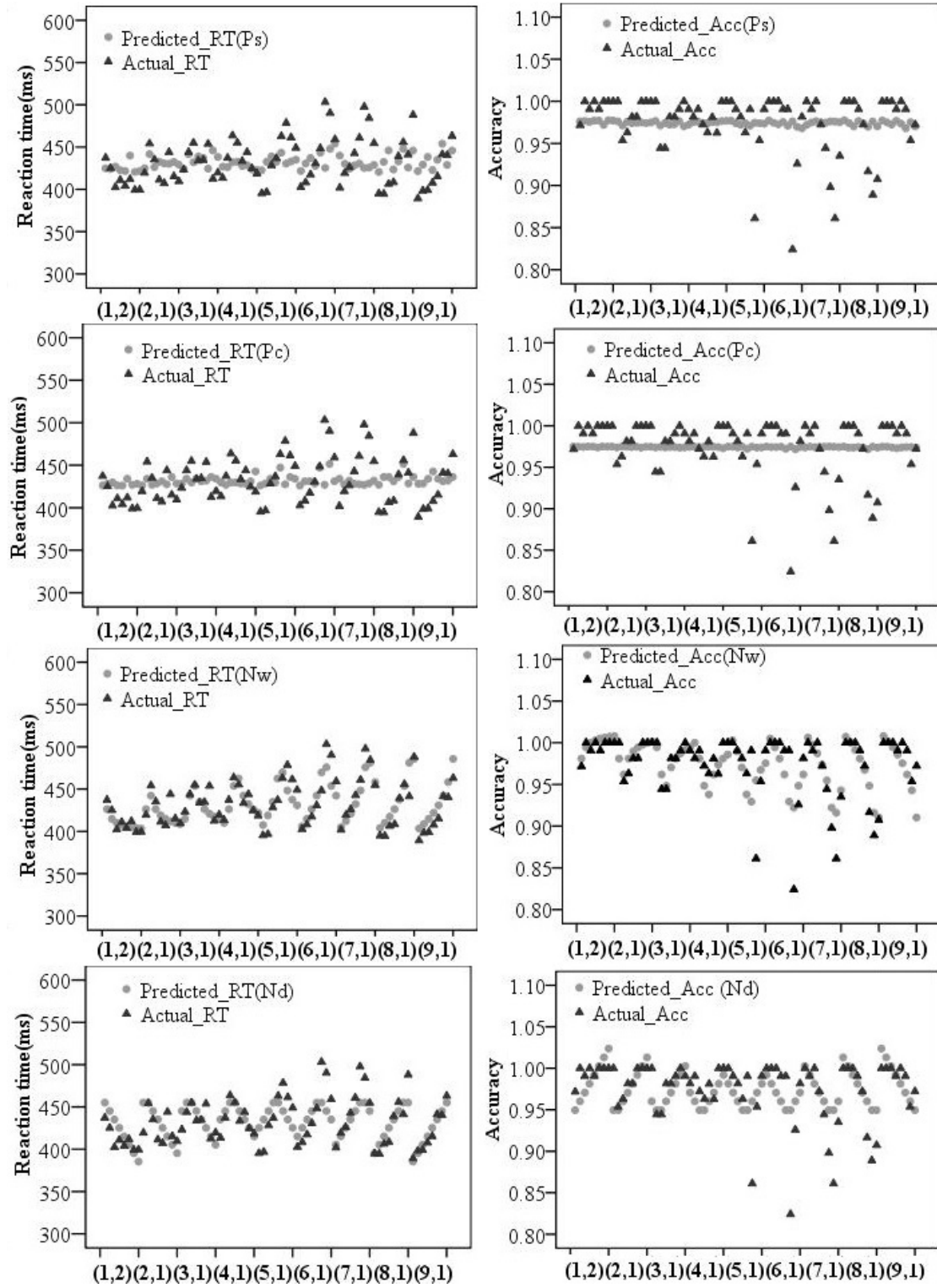




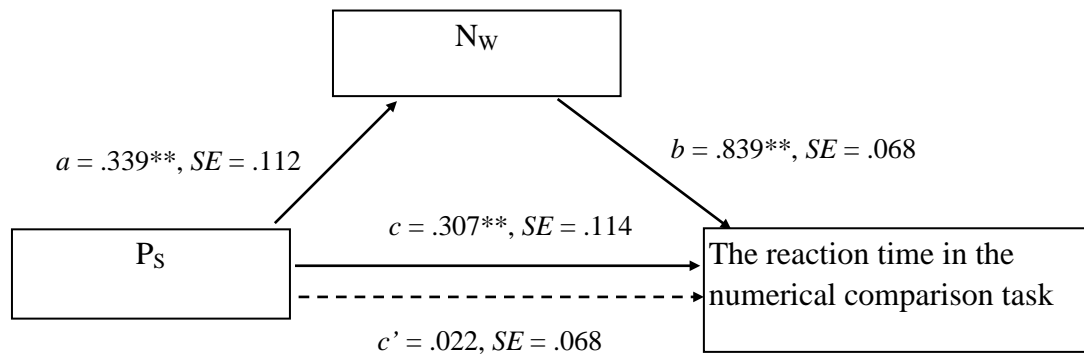
**Fig. 2.**  $P_S$  as a mediator between  $P_C$  and the reaction time for the simultaneous same-different task (Experiment 1). Path coefficients for mediation analysis on the reaction time. *Note:* Solid line denotes the total effect of  $P_C$  on the reaction time when  $P_S$  is included as a mediator. Dotted line denotes the direct effect of  $P_C$  on the reaction time when  $P_S$  is included as a mediator. Thus, the indirect effect of  $P_C$  on the reaction time is  $0.328(p < .001)$ .  $a$ ,  $b$ ,  $c$  and  $c'$  are unstandardized OLS regression coefficients. \*\*\* $p < .001$ .



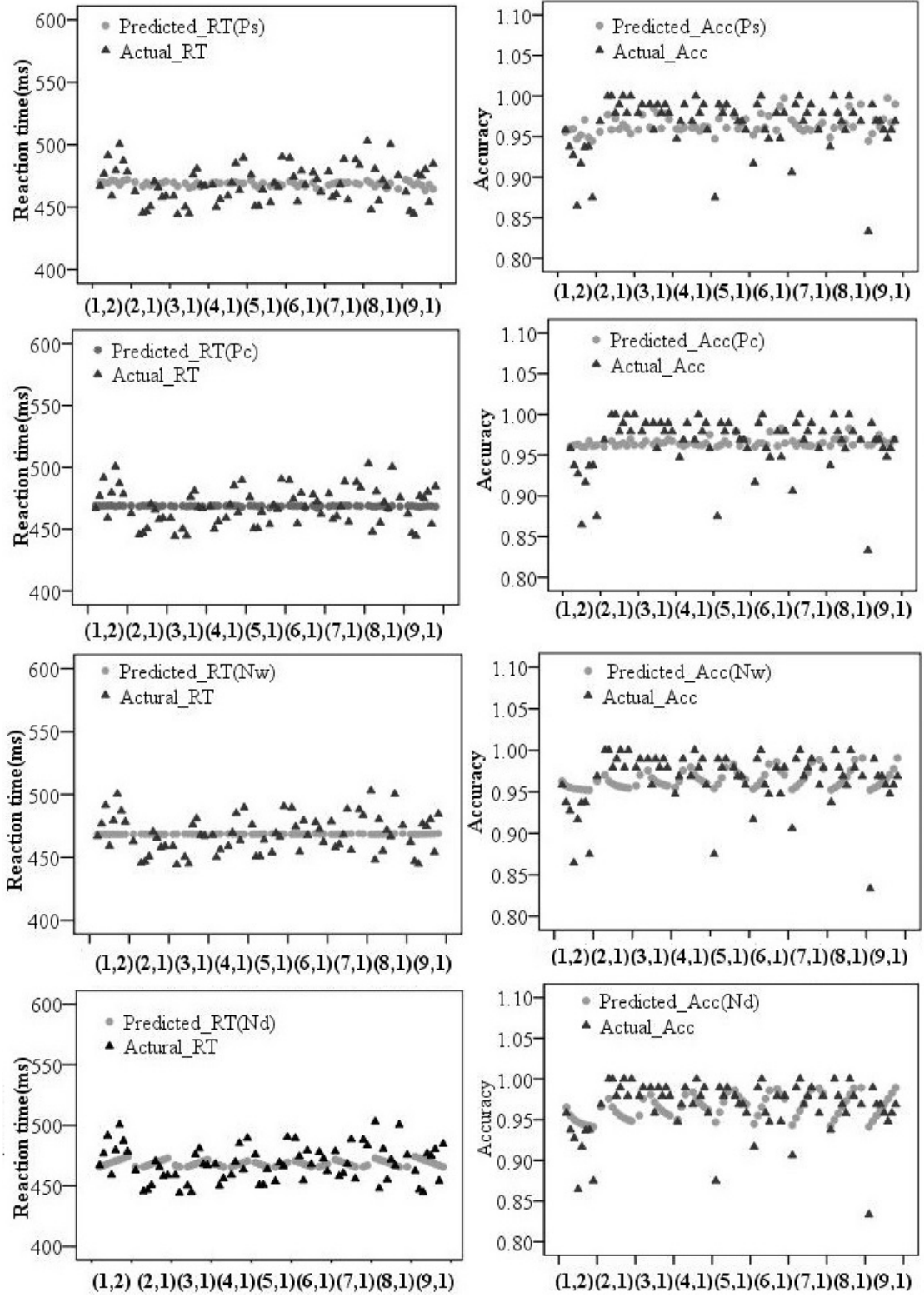
**Fig. 3.** Plot of actual average RT/Accuracy and predicted RT/Accuracy based on  $P_s$ ,  $P_c$ ,  $N_w$ ,  $N_d$  for the sequential same-different task (Experiment 1). Only  $P_s$  predicted the reaction time.



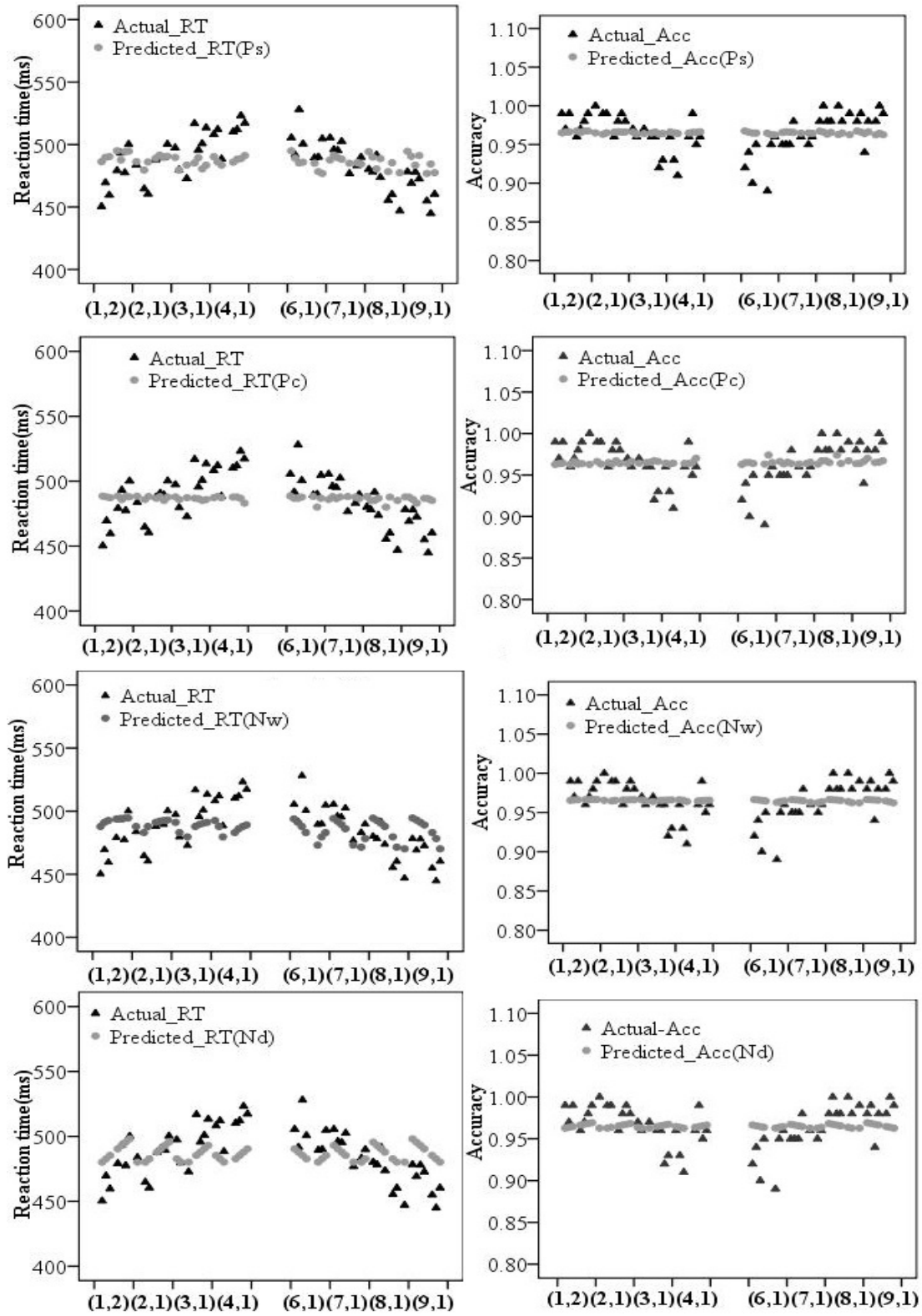
**Fig. 4.** Plot of actual average RT/Accuracy and predicted RT/Accuracy based on  $P_s$ ,  $P_c$ ,  $N_w$ ,  $N_d$  for the numerical comparison task (Experiment 2). Only  $N_w$  was a significant predictor of RT/Accuracy in simultaneous multiple linear regression.



**Fig. 5.**  $N_W$  as a mediator between  $P_S$  and the reaction time in the numerical comparison task (Experiment 2). The meaning of  $a$ ,  $b$ ,  $c$ ,  $c'$  is similar to Fig. 2. In Experiment 2, the indirect effect of  $P_S$  on the reaction time is significant ( $p = .003$ ).  $^{**} p < .05$ .



**Fig. 6.** Plot of actual average RT/Accuracy and predicted RT/Accuracy based on  $P_s$ ,  $P_c$ ,  $N_w$ ,  $N_d$  for the physical comparison task (Experiment 3).  $P_s$  and  $N_d$  were significant predictors of accuracy in simultaneous multiple linear regression.



**Fig. 7.** Plot of actual average RT/Accuracy and predicted RT/Accuracy based on  $P_s$ ,  $P_c$ ,  $N_w$ ,  $N_d$  for the priming task (Experiment 4). Both  $N_w$  and  $N_d$  significantly predicted the reaction time.

